Article

Current Biology

Global Patterns and Drivers of Bee Distribution

Highlights

- Bees show a rare bimodal latitudinal gradient with highest richness at mid-latitudes
- Xeric and temperate zones host higher richness than tropical areas
- Plant productivity and richness are important drivers when forests are excluded
- A global bee species richness reconstruction is presented for the first time

Authors

Michael C. Orr, Alice C. Hughes, Douglas Chesters, John Pickering, Chao-Dong Zhu, John S. Ascher

Correspondence

achughes@xtbg.ac.cn (A.C.H.),
dbsajs@nus.edu.sg (J.S.A.)

In Brief

A modern, quantitative synthesis on bee distribution and its drivers at a global scale. Orr et al. show that bees exhibit a rare bimodal pattern of higher species richness at mid-latitudes, based on their great success in xeric and some temperate areas, further supported by a driver analysis. Bee species richness is also reprojected worldwide.

Orr et al., 2021, Current Biology *31*, 1–8
February 8, 2021 © 2020 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2020.10.053

CellPress

Please cite this article in press as: Orr et al., Global Patterns and Drivers of Bee Distribution, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.053

Current Biology

CellPress
OPEN ACCESS

Article

Global Patterns and Drivers of Bee Distribution

Michael C. Orr,[1] Alice C. Hughes,[2,3,4,8,*] Douglas Chesters,[1] John Pickering,[5] Chao-Dong Zhu,[1,4,6] and John S. Ascher[7,*]

[1]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Beijing 100101, China
[2]Landscape Ecology Group, Centre for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Xishuangbanna, Yunnan 666303, China
[3]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla 666303, China
[4]University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing 10049, China
[5]University of Georgia, Athens, GA 30602, USA
[6]State Key Laboratory of Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China
[7]Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 117558, Singapore
[8]Lead Contact
*Correspondence: achughes@xtbg.ac.cn (A.C.H.), dbsajs@nus.edu.sg (J.S.A.)

https://doi.org/10.1016/j.cub.2020.10.053

SUMMARY

Insects are the focus of many recent studies suggesting population declines, but even invaluable pollination service providers such as bees lack a modern distributional synthesis. Here, we combine a uniquely comprehensive checklist of bee species distributions and >5,800,000 public bee occurrence records to describe global patterns of bee biodiversity. Publicly accessible records are sparse, especially from developing countries, and are frequently inaccurate throughout much of the world, consequently suggesting different biodiversity patterns from checklist data. Global analyses reveal hotspots of species richness, together generating a rare bimodal latitudinal richness gradient, and further analyses suggest that xeric areas, solar radiation, and non-forest plant productivity are among the most important global drivers of bee biodiversity. Together, our results provide a new baseline and best practices for studies on bees and other understudied invertebrates.

INTRODUCTION

Insects are reportedly declining at alarming rates worldwide, yet we do not understand even the most basic elements of their distributional dynamics.[1] Despite their importance, knowledge of insect biodiversity remains remarkably poor; the sheer number of species and the difficulty of identifying them preclude typical monitoring approaches, and the requisite funding is lacking.[2,3] Consequently, millions of museum specimens await identification or even formal description, remaining inaccessible to researchers.

Understanding insect distribution is key to evolutionary studies of origin and diversification, as well as ecological or conservation-oriented studies of how specific groups will respond to threats such as climate change or other human-induced phenomena.[4,5] In light of this, building and sharing our knowledge of insect distribution is one of the greatest, most important challenges that biologists and conservationists face, but the challenges of studying insects mandate the study of representative areas or specific groups.

As ecologically and economically invaluable pollinators, bees represent an ideal case study.[4,6–8] However, comprehensive analyses of bee distribution are nearly non-existent, with most focusing on limited regions[9–11] or site-based studies.[12,13] Well-known, eusocial bumblebees (*Bombus*) and the less-studied, solitary polyester bees (Colletinae) are exceptions.[14,15] However, these groups comprise <4% of described bee species (802/20,355 species[16]). Furthermore, *Bombus* dominate at higher latitudes and elevations, whereas Colletinae are more species rich in xeric areas, suggesting that neither alone can represent overall bee biodiversity.

Those few efforts that explore worldwide bee distribution are descriptive, reliant on comparisons between small, well-sampled areas such as Palm Springs and Riverside in California.[17] Nonetheless, some general patterns have been hypothesized: bee species richness is highest in relatively xeric areas while tropical environments, famed for extraordinary insect species richness, have few.[17] This leads to a bimodal latitudinal gradient in some bee groups.[15,18,19] However, data remain limited for testing reported global trends, and supposed bimodal latitudinal gradients of other Hymenoptera are uncertain due to sampling biases and taxonomic under-description,[20] leaving few documented examples.[21] To date, these hypotheses remain untested for bees globally. The primary cause of this bee distribution knowledge gap is insufficient reliable occurrence data,[22] although the analytical and taxonomic expertise required have also precluded exhaustive analysis of bee distribution.

Here, we map and model the known distribution of bees based on a uniquely comprehensive checklist collated from specimens, verified observations, and published records, and quantitatively compare this to occurrence data from five public databases.[16] In doing so, we reveal the biases of public bee occurrence data and provide best practices for future analyses. By combining multiple, mutually informative data sources, we generate the most

© 2020 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: Orr et al., Global Patterns and Drivers of Bee Distribution, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.053

Table 1. Public Data Filtering Results by Source

Data Source	Original	Duplicates	Taxonomy	Hemisphere	Species
IDB	1,973,815 (100)	223,338 (11.3)	216,300 (11)	205,265 (10.4)	9902
GBIF	1,514,040 (100)	389,261 (25.3)	384,573 (25)	371,643 (24)	8174
BISON	1,315,811 (100)	229,381 (17.4)	205,280 (15.6)	195,030 (14.8)	3388
SCAN	910,947 (100)	183,182 (20.1)	118,245 (13)	109,515 (12)	4574
ALA	116,198 (100)	30,622 (26.4)	26,741 (23)	25,548 (22)	919

This includes IDB (iDigBio), GBIF (Global Biodiversity Information Facility), BISON (Biodiversity Information Serving our Nation), SCAN (Symbiota Collections of Arthropods Network), and ALA (Atlas of Living Australia), arranged in descending order of number of original records following each step. Percentages of original records are listed in brackets, with 100% listing the entire dataset, and subsequent values left following that filter step: following duplicate removal, following synonym check, and following hemispheric check, respectively. Table S2 contains the synonym filter. See also Table S1 and Data S1.

comprehensive assessment of global bee distribution, delimiting world hotspots of bee species richness. We then assess the drivers of these patterns and, in turn, use these predictions to model bee richness worldwide.

RESULTS

Public Database Cleaning and Comparison to Checklist

Of the 5,857,811 occurrence records compiled, under 16% (907,001) passed all filters (Table 1). When excluding duplicate removal steps, which are not indicative of error and constituted 75%–90% of records (and also included duplicates between the databases; Table S1), there is an overall error rate span of 1%–8% for the datasets. 43,857 synonyms were compiled for the world total of 20,555 bee species, and 10,724 records were corrected across 6,340 species (Table S2). Although the East-West hemispheric check detected negligible error rates (<1%), this translates to 1,703 species in the wrong hemisphere.

Further checklist-based validations on the database at the country-level revealed many more errors. The average percent of incorrect species was 10% across all 25 checked countries, with a maximum of 19.4% (Malaysia) and a minimum of 0% (Philippines) (10,358 records). All checked countries that originally had more putative species than the checklist subsequently had fewer species than the checklist when erroneous species were removed (Figure S1; Methods S1A). Although only 1% of the cleaned USA samples are incorrectly recorded, this is 10.4% (377/3,435) of total species, showing that incorrect singletons (72% of incorrect USA records) are why some areas have more putative species listed in the public database than in the checklist (Figure S1; Data S1).

Patterns in the public database differed profoundly from the checklist, recovering differing richness hotspots and radically lower richness in developing regions, particularly evident when using cartograms that distort areas depending on their relative over- or under-sampling (Figure 1). Richness across much of Asia and Africa (except South Africa) is dramatically lower according to the public data, evidentially a result of low sampling effort combined with insufficient data sharing in some regions (Figure 1B). For example, the USA represents >60% of non-duplicated public records, more than the rest of the world combined, and the state of California has more than double the number of records of any country (except Sweden). Contrastingly, the best-sampled countries in Africa, Asia, and South America represent only 2% (South Africa), 0.25% (Japan), and 0.89% (Brazil) of all records, respectively. There is even a marked disparity between species richness and sampling intensity regionally; Africa, Latin America, Asia, and the Middle East contain 84% of global species yet only represent 12% of cleaned samples (Figure 2).

As a consequence of public data gaps and biases, many countries have exceedingly low spatial coverage and completeness compared to the checklist (Figure 3; Data S1A−S1G). Almost 15% of countries have under 5% of their land area sampled, and 55% have under 25% of their area sampled, whereas only 12% of countries have over 95% of their area sampled. This equates to 90% of North Asia, 86% of South Asia, 82% of Middle East, and 79% of Africa and Eurasia completely unsampled (Data S1). Some of the most species-rich countries have the least available data; for example, China has the sixth-most bee species (6%) but records exist for only 7% of species, averaging two records per species listed, with only 0.03% of global cleaned records. In contrast, the USA contains >60% of databased samples and 17.5% of recorded species (113 samples per species). The second most species-rich country is Mexico, with 9.1% of global species and 3% of samples (14 samples per species). In third place is Brazil at 8.9% of total species, but with only 0.9% of samples and six records per species. Country-level rarefactions based on public data suggest similar biases, with estimated richness for only the UK and USA exceeding 90% of checklist richness following initial checks (Figure S2; Methods S1B).

Bee Species Richness Worldwide

The checklist provides a clearer picture of bee distribution, in light of pervasive public data biases. Large hotspots of richness are apparent in the southwestern USA, Mediterranean Basin into the Middle East, and Australia, with a weaker signal in South Africa (Figures 1 and 4). Israel has the highest richness-per-unit area (when removing areas under 5,000 km^2), though the USA (especially western states), the Mediterranean, Nepal, areas around the Andes, areas south of the Amazon Basin in South America, and South Africa also have high levels of area-weighted richness (Data S1). Contrasting with species-rich arid-temperate areas, the humid tropics and even arid-tropical areas are generally much poorer.

The minimum convex polygon stacking visualization of the public database highlights the hotspot in the southwestern USA, with decreasing richness toward the tropics (Figures 1

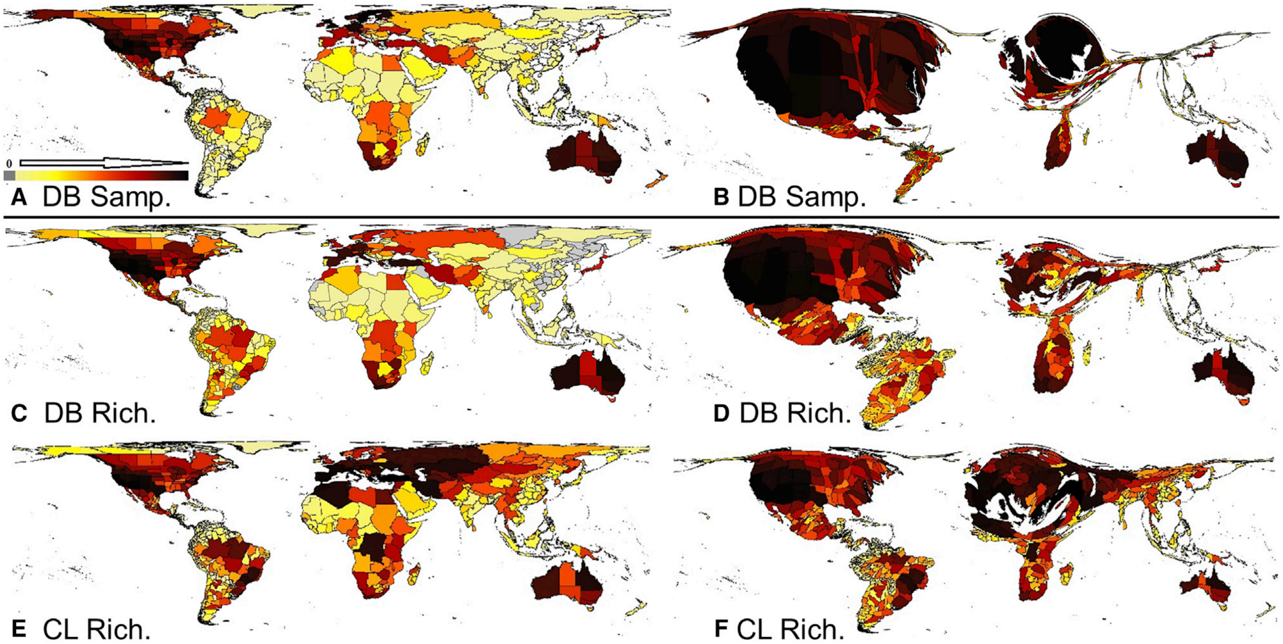

Figure 1. Patterns of Bee Distribution

Maps (left) and cartograms (right) are given for (A and B) number of public database records post-cleaning, (C and D) patterns of species richness in the cleaned public database, and (E and F) patterns of richness in the checklist data. Darker areas have higher values. In general, the public dataset strongly followed sampling effort, while the checklist recovered more representative patterns. Scale given in (A) applies to all. Data S1 includes country-level totals for reference.

and 5A) and peaks in species range limits at the tropical-subtropical interface (Figure S3A). Looking at faunal similarity, the sPCA suggests that the temperate eastern and central USA are relatively similar but that the western hot and cold deserts are quite distinctive, extending to Central Mexico, there transforming into the less-rich Central and South American fauna (Figure 5B). South America appears to have a more distinctive tropical fauna centered around the Amazon Basin, various elements of which extend far outward into Central and throughout northern South America. The turnover analysis recovers high turnover along coastal areas generally, but in North America, distinct Atlantic and Gulf faunas are suggested, as well as the xeric Southwest and adjacent areas, down through much of Central America (Figure 5C). There is generally less turnover in the Amazon Basin, though higher in northern Brazil, southeastern Brazil, Chile, and generally west of the Andes. All three models fail to reconstruct patterns in southern South America, likely due to the combination of poorer sampling and high endemicity (Figure 1C).[23]

The richness peaks at mid-latitudes in both the Northern (30°–40°) and Southern Hemispheres (–30°) from the checklist clearly affirm a bimodal latitudinal gradient for bees (Figures S4A and S4B). Bimodal peaks are evident for the New World (somewhat weakly), Europe-Africa, and Australasia (Figure 4). Hotspots in the Northern Hemisphere have much higher richness per-unit-area than in the Southern Hemisphere, contributing to the lessening of the bimodal latitudinal gradient in the South when accounting for area (Figure 4). Weighting richness by area does not change the latitudinal gradients substantially (Figure S4A), but reduces the magnitude in the South, likely in part due to smaller land area.

Drivers of Bee Richness

Of 62 variables hypothesized to influence global bee distribution, 24 had significant relationships with checklist species richness in independent regressions. Multiple approaches were used to explore these relationships, but all showed similar variables to

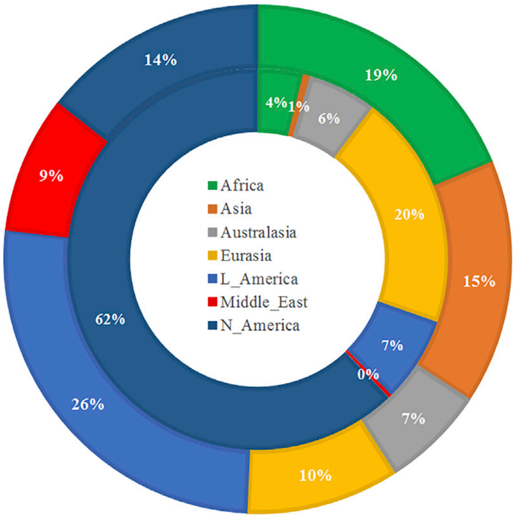

Figure 2. Sample Number and Expected Richness by Region

Percentage of samples in the public database is given on the interior ring, and percentage of total global bee species richness based on the checklist is given on the exterior ring. The vast majority of records comes from areas (North America and Eurasia) with a minority of global species. For listing and a map of the countries that fall into each region, please see Data S1.

Please cite this article in press as: Orr et al., Global Patterns and Drivers of Bee Distribution, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.053

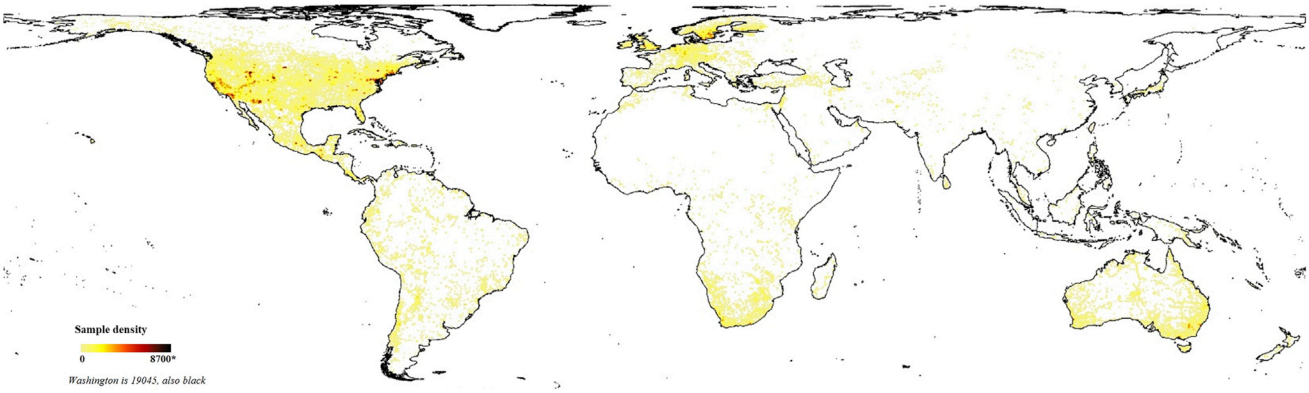

Figure 3. Public Database Spatial Coverage and Sampling Density

Given on a 25-km^2 grid. Transparent areas are unsampled, while darker colors indicate more intense sampling. Sampling effort is clearly higher in developed countries, although even in these areas, the most-sampled areas are highly localized. Additional error checks and supporting data are in Figures S1 and S2, Tables S1 and S2, and Data S1.

be of the greatest importance (Data S2–S4; Methods S1C). Variation in solar radiation, mean seasonality in potential evapotranspiration, and mean continentality showed some of the strongest relationships.

The final model included 18 well-supported, independent variables, with 759 administrative areas under consideration ($r = 0.775$, r^2 Adj = 0.592, AICc = 9602.895; Methods S1C). The Geoda model assessed collinearity and reduced it to below 30 and r^2 fell, but it provided similar outputs (Figure S4C; Data S2 and S3; Methods S1C). Major drivers included several components of solar radiation, showing that high solar radiation (available energy) and lower variation correspond to high bee richness. The analysis supports the view that bee richness is highest in areas with high solar insolation, as expected given benefits to plant growth and bee thermoregulation;[24] but sufficient moisture for plant growth is also needed, and thus high potential evapotranspiration mean is important. Low levels of precipitation during the driest month and reduced seasonality also enhance high richness, supporting the view that deserts are important areas for bees, in addition to Mediterranean climates (as shown by the positive relationship with Embergers Pluviothermic quotient and Figures S3B–S3D). Conversely, more growing degree days and lower wind were beneficial, corresponding to polar and higher-elevation areas holding generally lower bee richness. High net primary productivity (minus forests) also correlated with higher bee richness. However, when forests were included, this relationship reversed. The three-region driver analysis largely recovered different most-important parameters for each area, clearly demonstrating that drivers vary by region and scale, although max temperature of warmest month and plant species richness (without forests) were consistently important across all three (Methods S1C).

The combined driver analysis and point data enabled a higher-resolution (10 km) view of relative global bee species richness

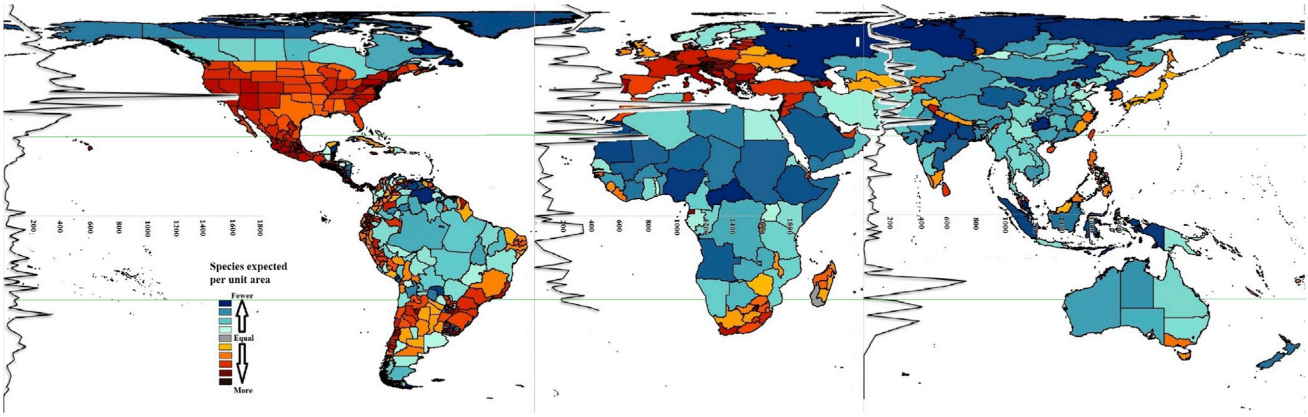

Figure 4. The Bimodal Latitudinal Gradient in Bees

Based on checklist data. Separate line graph gradients displaying absolute species richness trends are given for the New World, Europe-Africa, and Australasia. Conversely, showing a different metric, administrative polygons are colored according to cartogram-derived difference from expected richness given the size of each unit, with darker reds being higher than expected richness and darker blues being lower than expected; thus, areas with more species than the global average per unit area are in amber-red, and those with fewer species per unit area are in cyan-navy. Clear richness peaks are evident in the Northern and Southern Hemispheres while lacking in the tropics, both when controlling for area and not. Green lines indicate the boundary of the tropics, with the area between the two considered as tropical. This is supported by Figure S3A, which highlights peaks in range limits on the American continent.

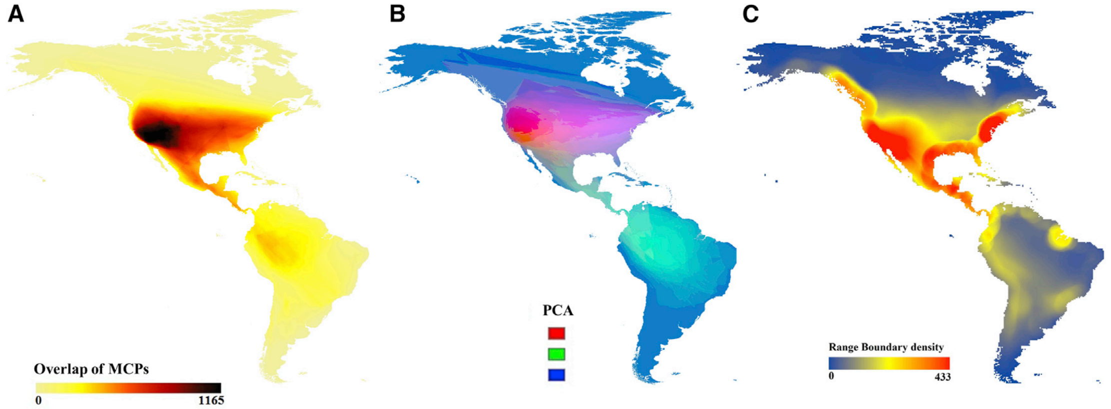

Figure 5. Minimum Convex Polygon Mapping across New World Bees

Based on public database. Shown are (A) richness of polygons, (B) sPCA, to show community composition and changes, and (C) turnover (based on the number of range boundaries). All three methods suggest a large, distinct southwestern USA fauna, but sampling limitations hinder these reconstructions of the South American fauna. See Figure S3A for the graph of maximum and minimum range limits.

patterns (Figure 6; Methods S1D). The New World model (using global drivers) was projected for the Old World due to inadequate Old World data, and consequently it is less applicable to environments truly unique to the Old World (the Old World models based on Old World data are shown in Figure S5), where more data will be necessary to fully understand bee richness patterns at finer scales. Although more work is needed in the Old World tropics, there is no evidence suggesting a bounty of undescribed species there when compared to under-described areas such as in China or Australia, such that the bimodal pattern would hold or become even stronger; as our study uses relative richness, these issues are largely already accounted for. Additionally, future approaches will need to better account for island communities, as many were overinflated in the present model because they are based solely on environmental potential to host species rather than biogeographic limitations, which prevent migration or potential resource limitations (specific soils for nesting, floral resources, etc.). Nonetheless, Mediterranean and xeric areas are recognized for their high richness as in other approaches, with some notably high temperate areas such as parts of the northern USA, southern South America, some of South Africa, northern China, and the Himalayan foothills, and the bimodal latitudinal gradient is again strongly supported.

DISCUSSION

Distributional information shapes our knowledge of species, directly informing both conservation and management decisions locally and regionally. Unfortunately for bees and most other invertebrates, the coverage and quality of these data are severely lacking despite calls for data mobilization.[25] Reported insect declines make this work imperative,[26–28] but conflicting accounts underscore the need for more thorough analyses.[29] Although sampling and digitization biases are unavoidable, they are severe enough here that they can effectively obscure true biodiversity patterns when combined with data quality issues (Figure 1).

Going forward, data cleaning and clear standards must be recognized as equally important to data generation. Although this may limit data generation, improved relations and resourcing between museum specialists and end data users may mitigate increased workloads when combined with sufficient funding and increased recognition for primary data generation.[30] All analyses suggest that while regions such as parts of North America and northern Europe are well known, Africa and Asia are not (Figure 2; Figure S4C), and knowledge of the Australian fauna is largely limited to coastal areas (Figure 3) despite apparent high richness in less-accessible regions.[31] Ultimately, processing, curation, and digitization of museum specimens from less-known areas should be viewed as more important than the inventories and expeditions from which they come, as data are only useful when reliable, accurate, and accessible.

In combination, our checklist and cleaning processes enable a more accurate and detailed view of bee distribution than ever before, empirically supporting prior hypotheses and refining others.[17,32] Both public and checklist data show that the Northern Hemisphere clearly holds higher described species richness. This agrees with a smaller-scale study by Moldenke,[23] which found far fewer species in Mediterranean Chile than in Mediterranean California, but until now it was unknown whether this held worldwide. Overall, xeric-temperate areas outperform other regions in bee species richness while tropical areas underperformed, supporting prior hypotheses.[10,17,32] Higher-resolution analyses echoed these patterns (Figure 6) while also highlighting the lack of thorough and fully accessible species inventories outside North America, underscoring the need for greater digitization and data sharing.

Several temperate areas also appear to be unusually species rich, which is less expected[17] (Figures 1 and 6); this expands upon an emerging notion of Michener[32] (for southeastern Brazil) by identifying such areas worldwide. It is unclear what makes certain temperate areas more species rich, but distinct, overlapping faunas may play a role. For example, São Paulo state in Brazil resides at the interface of the neotropical with southern South American faunas undetected by the sPCA but more evident in the turnover analysis (Figure 5). Institutional proximity may also play a role; strong historical bee programs in Illinois, Washington, D.C., New York, and North Carolina all coincide with higher sampling and checklist richness (Figure 1).

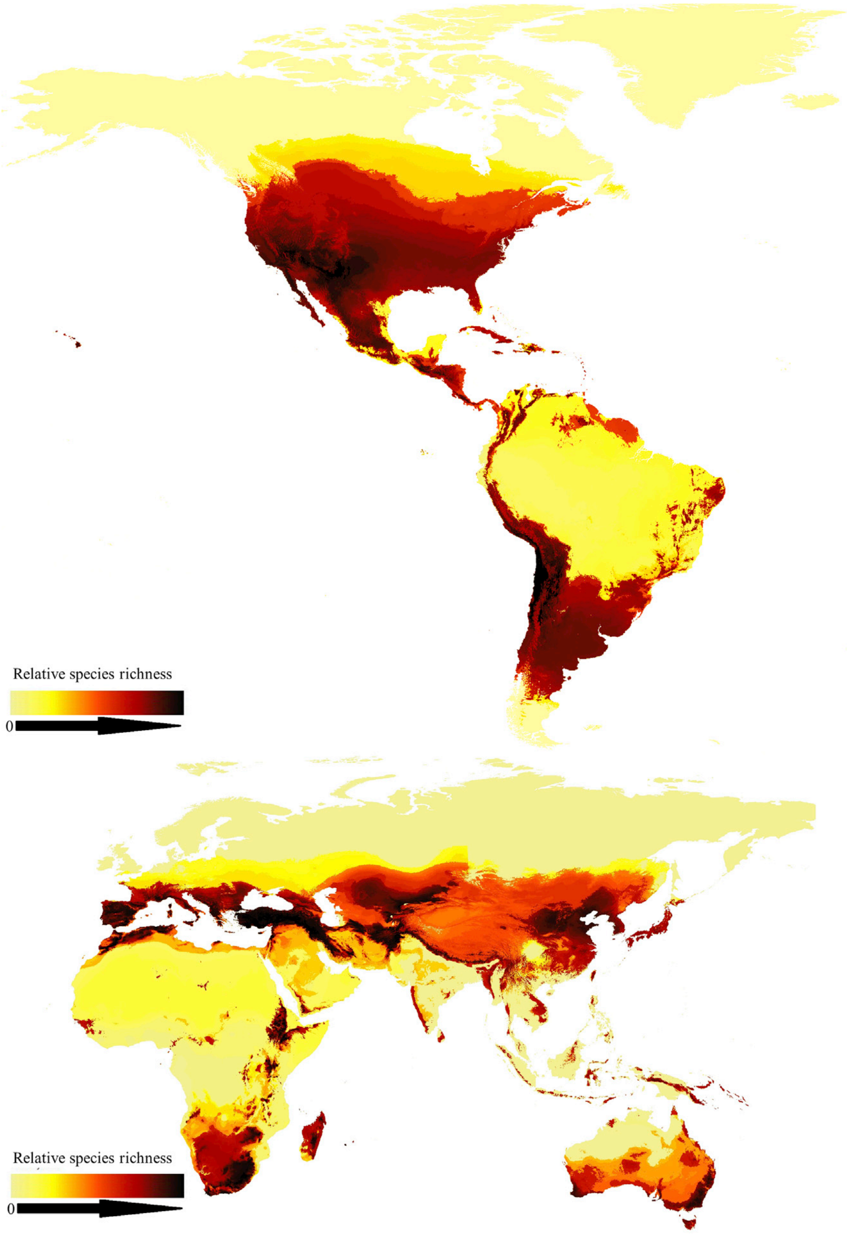

Figure 6. High-Resolution Bee Species Richness Projections

Checklist and point data were used (Methods S1A), with driver components extrapolated from New World to Old World given better sampling, which may limit its applicability in unique environments. Areas of higher projected richness are darker, but values are relative rather than absolute. Areas of lowest richness are in some cases underestimated due to insufficient data in comparable regions, while islands are generally overestimated (and Oceanic islands were largely removed from analysis). Bottom-up or hierarchical analyses will improve these models in the future. Further model details are available in Figures S3–S5, Table S3, Data S2–S4, and Methods S1.

The global patterns outlined here appear to be largely driven by energy (solar) and resource (water and plants) availability within a relatively less-stringent climatic envelope (Methods S1C). These factors were more consistently important across analyses, while other factors often changed depending on whether global or regional scales were analyzed. Given these patterns of richness and their reliance on some climatic factors, global climate change, especially fluctuating seasonality and subsequent impacts on plant phenology, could impact bees in complex ways, but additional analysis is necessary to explore how. Most drivers supported the highest bee species richness at relatively intermediate values when considering their potential maximums and minimums (e.g., a place with no precipitation would be unfavorable even if bees prefer drier climates, as they require at least some minimum threshold).

Future models may be improved by incorporating finer-scale driver analyses, such as via ecoregional classifications or point data when available (Figure 3). Similarly, finer-scale experiments will be necessary to understand the proximal drivers of bee species richness, including accounting for specific habitat types, but the general negative relationship between water and bee richness suggest that humidity may play a key role in limiting bee distribution (such as through spoilage of pollen resources for solitary bees).

The relationship between bees and net primary productivity (minus forests) (and plant species richness regionally) is especially interesting, as bee distribution should intuitively be related to flowering plants, but such a linkage had not yet been established at larger scales. This is likely because straightforward metrics of plant richness, even controlling for forestation as we did here, do not correlate strongly with bee richness except at regional resolution (Data S3E; Methods S1C). It seems likely that future studies using additional taxonomic or functional subdivisions of bees and plants will reveal similar relationships, although they may be more nuanced. For instance, whether trees might prove more beneficial to bee species richness in tropical areas where more trees provide floral resources than in other environments requires further, finer-scale study of different habitat types within the tropics, but globally it is clear that bee richness is negatively impacted by trees.

These drivers and likely the biogeographic history of the bees together generate a bimodal latitudinal gradient in bees, supporting prior hypotheses (Figure 4). This strongly contrasts with other pollinator groups and many other taxa, which typically achieve their greatest richness near the equatorial tropics.[10,21] This is not an artifact of under-description in the tropics: the percentage of new species in heavily sampled inventories in the species-rich, xeric areas of North America (11% [48/450 total] undescribed species in Pinnacles National Park[33] and 7%

[49/660] in Grand Staircase-Escalante National Monument[34]) are similar to or slightly lower than those from tropical areas (12% [42/353] undescribed in Panama,[35] <16% [20/130] in a Belize forest,[36] and <20% [25/127] in Singapore[37]). Considering total species numbers, this description disparity is clearly insufficient to negate the latitudinal patterns seen here, and projected richness patterns that better account for undescribed species reinforce these patterns (Figure S4C).

This study outlines bee richness globally, but many questions remain. More representative point locality data will greatly improve the resolution and depth of our knowledge, enabling more powerful analyses and knowledge of how bees interact with different environments. Well over a century of life history data exist for bees, and these can be combined with distributional information to reveal generalizable patterns of where traits such as cavity nesting or floral specialization are more prevalent. Naturally, different groups will thereby show disparate patterns (e.g., highly eusocial honeybees and stingless bees are more prevalent in tropical areas[32]). Although difficult, such complexities must be accounted for to understand and map the history of bee evolution.

Ongoing targeted data-capture efforts can only improve our understanding of insect richness, but waiting for the digitization of all specimens may take decades, and only one insect group so far has been IUCN assessed, though most vertebrates have been.[28,38] Funding, personnel, and expertise are obvious limitations on digitization,[11] and given that the current model of academia does not properly reward data generation and maintenance, institutional infrastructure will be necessary.[39]

A well-funded, singular data repository could contract experts to build and share similar checklists. This would minimize errors via active checklist validations and reports to data owners, enable georeferenced cross-checking across taxa for errors, easily automate the elimination of duplicates, correct synonyms, provide sensibly formatted and easy-to-download access options, and resolve many issues inherent to reconciling multiple data formats. In these ways, both providing and using data would become far easier, which should greatly increase participation and value of the data for research and management. These measures, as demonstrated here, would greatly improve our ability to understand the natural world.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead Contact
  - ○ Materials Availability
  - ○ Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ Data compilation and cleaning
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Mapping and data visualization
  - ○ Rarefaction analyses
  - ○ Drivers analyses and mapping richness
  - ○ Regional diversity patterns

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.10.053.

ACKNOWLEDGMENTS

We thank Yanhua Chen for assistance collating synonyms, and Xin Rui Ong is thanked for copyediting. Paul H. Williams, Jeff Ollerton, Pierre Rasmont, Amber D. Tripodi, Vasuki Belavadi, Rafael Ferrari, Maxim Y. Proshchalykin, and Natapot Warrit are thanked for comments. Roman Carrasco and Catharina Karlsson are thanked for input on the statistics. The editor and four reviewers are thanked for their improvements. Harold Ikerd and Terry Griswold are thanked for discussions. M.C.O., D.C., and C.-D.Z. were mainly supported by the Strategic Priority Research Program of the Chinese Academy of Science (CAS) (XDB310304) and the National Science Fund for Distinguished Young Scholars (31625024). Other support included the Research Fund for International Young Scientists NSFC (31850410464) and the CAS PIFI (2018PB0003 and 2020PB0142) (to M.C.O.); the Strategic Priority Research Program of the CAS grant XDA20050202, the CAS Southeast Asia Biodiversity Research Center fund grant Y4ZK111B01, and the SEABA: Southeast Asian Atlas of Biodiversity grant Y5ZK121B01 (to A.C.H.); and NSFC 31772495 (to D.C.); as well as the China-Singapore joint grant 41761144068/R-154-000-A81-281 (to C.-D.Z. and J.S.A.) and Singapore Ministry of Education AcRF grant R-154-000-A36-114 (to J.S.A.).

AUTHOR CONTRIBUTIONS

M.C.O., A.C.H., and J.S.A. conceived the study. J.S.A. compiled the checklist data and J.P. developed error-checking and display tools. A.C.H. and M.C.O. aggregated the public data and A.C.H. cleaned these data and conducted primary analysis. A.C.H., D.C., and M.C.O. conducted secondary analyses. M.C.O., J.S.A., and A.C.H. interpreted the results. M.C.O. and A.C.H. wrote the initial draft. All authors contributed to and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 12, 2020
Revised: August 27, 2020
Accepted: October 19, 2020
Published: November 19, 2020

REFERENCES

1. Stork, N.E. (2018). How many species of insects and other terrestrial arthropods are there on Earth? Annu. Rev. Entomol. *63*, 31–45.
2. Clark, J.A., and May, R.M. (2002). Taxonomic bias in conservation research. Science *297*, 191–192.
3. Leather, S.R. (2013). Institutional vertebratism hampers insect conservation generally, not just saproxylic beetle conservation. Anim. Conserv. *16*, 379–380.
4. Potts, S.G., Imperatriz-Fonseca, V., Ngo, H.T., Aizen, M.A., Biesmeijer, J.C., Breeze, T.D., Dicks, L.V., Garibaldi, L.A., Hill, R., Settele, J., and Vanbergen, A.J. (2016). Safeguarding pollinators and their values to human well-being. Nature *540*, 220–229.
5. Settele, J., Bishop, J., and Potts, S.G. (2016). Climate change impacts on pollination. Nat. Plants *2*, https://doi.org/10.1038/nplants.2016.92.
6. Klein, A.M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., and Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. Proc. Biol. Sci. *274*, 303–313.
7. Scheper, J., Reemer, M., van Kats, R., Ozinga, W.A., van der Linden, G.T., Schaminée, J.H., Siepel, H., and Kleijn, D. (2014). Museum specimens reveal loss of pollen host plants as key factor driving wild bee decline in The Netherlands. Proc. Natl. Acad. Sci. USA *111*, 17552–17557.

8. Powney, G.D., Carvell, C., Edwards, M., Morris, R.K.A., Roy, H.E., Woodcock, B.A., and Isaac, N.J.B. (2019). Widespread losses of pollinating insects in Britain. Nat. Commun. *10*, https://doi.org/10.1038/s41467-019-08974-9.

9. Nieto, A., Roberts, S.P.M., Kemp, J., Rasmont, P., Kuhlmann, M., García Criado, M., Biesmeijer, J.C., Bogusch, P., Dathe, H.H., De la Rúa, P., et al. (2014). European red list of bees (International Union for Conservation of Nature, Luxembourg, Publication Office of the European Union).

10. Ollerton, J. (2017). Pollinator diversity: distribution, ecological function, and conservation. Annu. Rev. Ecol. Evol. Syst. *48*, 353–376.

11. Bartomeus, I., Stavert, J.R., Ward, D., and Aguado, O. (2018). Historical collections as a tool for assessing the global pollination crisis. Philos. T. R. Soc. B. *374*, https://doi.org/10.1098/rstb.2017.0389.

12. Archer, C.R., Pirk, C.W.W., Carvalheiro, L.G., and Nicolson, S.W. (2014). Economic and ecological implications of geographic bias in pollinator ecology in the light of pollinator declines. Oikos *123*, 401–407.

13. De Palma, A., Abrahamczyk, S., Aizen, M.A., Albrecht, M., Basset, Y., Bates, A., Blake, R.J., Boutin, C., Bugter, R., Connop, S., et al. (2016). Predicting bee community responses to land-use changes: Effects of geographic and taxonomic biases. Sci. Rep. *6*, https://doi.org/10.1038/srep31153.

14. Williams, P. (2007). The distribution of bumblebee colour patterns worldwide: possible significance for thermoregulation, crypsis, and warning mimicry. Biol. J. Linn. Soc. Lond. *92*, 97–118.

15. Bystriakova, N., Griswold, T., Ascher, J.S., and Kuhlmann, M. (2018). Key environmental determinants of global and regional richness and endemism patterns for a wild bee subfamily. Biodivers. Conserv. *27*, 287–309.

16. Ascher, J.S., and Pickering, J. (2018). Discover Life bee species guide and world checklist, (Hymenoptera: Apoidea: Anthophila), Draft 51. https://www.discoverlife.org/mp/20q?guide=Apoidea_species.

17. Michener, C.D. (1979). Biogeography of the bees. Ann. Mo. Bot. Gard. *66*, 277–347.

18. Wcislo, W.T. (1987). The roles of seasonality, host synchrony, and behavior in the evolutions and distributions of nest parasites in Hymenoptera, (Insecta), with special reference to bees, (Apoidea). Biol. Rev. Camb. Philos. Soc. *62*, 515–542.

19. Petanidou, T., Ellis, W.N., and Ellis-Adam, A.C. (1995). Ecogeographical patterns in the incidence of brood parasitism in bees. Biol. J. Linn. Soc. Lond. *55*, 261–272.

20. Quicke, D.L.J. (2012). We know too little about parasitoid wasp distributions to draw any conclusions about latitudinal trends in species richness, body size and biology. PLoS ONE *7*, e32101.

21. Kindlmann, P., and Dixon, A.F.G. (2007). Inverse latitudinal gradients in species diversity. In Scaling Biodiversity, D. Storch, P.A. Marquet, and J.H. Brown, eds. (Cambridge: Cambridge University Press), pp. 246–257.

22. Isaac, N.J., and Pocock, M.J. (2015). Bias and information in biological records. Biol. J. Linn. Soc. Lond. *115*, 522–531.

23. Moldenke, A.R. (1976). Evolutionary history and diversity of the bee faunas of Chile and Pacific North America. The Wasmann Journal of Biology *34*, 147–178.

24. Willmer, P.G., and Stone, G.N. (2004). Behavioral, ecological, and physiological determinants of the activity patterns of bees. Advances in the Study of Behavior *34*, 347–466.

25. National Research Council (2007). Status of Pollinators in North America (Washington, D.C.: The National Academies Press).

26. Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., et al. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE *12*, e0185809.

27. Lister, B.C., and Garcia, A. (2018). Climate-driven declines in arthropod abundance restructure a rainforest food web. Proc. Natl. Acad. Sci. USA *115*, E10397–E10406.

28. Eisenhauer, N., Bonn, A., and A Guerra, C. (2019). Recognizing the quiet extinction of invertebrates. Nat. Commun. *10*, https://doi.org/10.1038/s41467-018-07916-1.

29. Willig, M.R., Woolbright, L., Presley, S.J., Schowalter, T.D., Waide, R.B., Heartsill Scalley, T., Zimmerman, J.K., González, G., and Lugo, A.E. (2019). Populations are not declining and food webs are not collapsing at the Luquillo Experimental Forest. Proc. Natl. Acad. Sci. USA *116*, 12143–12144.

30. Ward, D.F., Leschen, R.A.B., and Buckley, T.R. (2015). More from ecologists to support natural history museums. Trends Ecol. Evol. *30*, 373–374.

31. Leijs, R., Dorey, J., and Hogendoorn, K. (2018). Twenty six new species of *Leioproctus* (*Colletellus*): Australian Neopasiphaeinae, all but one with two submarginal cells (Hymenoptera, Colletidae, *Leioproctus*). ZooKeys *811*, 109–168.

32. Michener, C.D. (2007). The Bees of the World, Second Edition (Baltimore: Johns Hopkins University Press).

33. Meiners, J.M., Griswold, T.L., and Carril, O.M. (2019). Decades of native bee biodiversity surveys at Pinnacles National Park highlight the importance of monitoring natural areas over time. PLoS ONE *14*, e0207566.

34. Carril, O.M., Griswold, T., Haefner, J., and Wilson, J.S. (2018). Wild bees of Grand Staircase-Escalante National Monument: richness, abundance, and spatio-temporal beta-diversity. PeerJ *6*, e5867.

35. Michener, C.D. (1954). Bees of Panama. Bulletin of the American Museum of Natural History *104*, 1–176.

36. Bridgewater, S. (2012). A natural history of Belize: Inside the Maya Forest (University of Texas Press).

37. Soh, E.J.Y., Soh, Z.W.W., Chui, S.X., and Ascher, J.S. (2016). The bee tribe Anthidiini in Singapore, (Anthophila: Megachilidae: Anthidiini), with notes on the regional fauna. Nature in Singapore *9*, 48–62.

38. Clausnitzer, V., Kalkman, V.J., Ram, M., Collen, B., Baillie, J.E.M., Bedjanič, M., Darwall, W.R.T., Dijkstra, K.-D.B., Dow, R., Hawking, J., et al. (2009). Odonata enter the biodiversity crisis debate: the first global assessment of an insect group. Biol. Conserv. *142*, 1864–1869.

39. Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., and Frame, M. (2011). Data sharing by scientists: practices and perceptions. PLoS ONE *6*, e21101.

40. Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001). PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica *4*, 9.

41. Rangel, T.F., Diniz-Filho, J.A.F., and Bini, L.M. (2010). SAM: a comprehensive application for Spatial Analysis in Macroecology. Ecography *33*, 46–50.

42. Hsieh, T.C., Ma, K.H., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. *7*, 1451–1456.

43. Global Biodiversity Information Facility. https://www.gbif.org/.

44. Integrated DigBiol. http://www.idigbio.org.

45. Southwest Collections of Arthropods Network. https://www.scan-bugs.org/portal/index.php.

46. Atlas of Living Australia. https://www.ala.org.au.

47. Biodiversity Information Serving Our Nation. https://bison.usgs.gov.

48. Hoskins, A.J., Harwood, T.D., Ware, C., Williams, K.J., Perry, J.J., Ota, N., Croft, J.R., Yeates, D.K., Jetz, W., Golebiewski, M., et al. (2019). Supporting global biodiversity assessment through high-resolution macroecological modelling: Methodological underpinnings of the BILBI framework. bioRxiv. https://doi.org/10.1101/309377.

49. Mokany, K., Ferrier, S., Harwood, T.D., Ware, C., Di Marco, M., Grantham, H.S., Venter, O., Hoskins, A.J., and Watson, J.E.M. (2020). Reconciling global priorities for conserving biodiversity habitat. Proc. Natl. Acad. Sci. USA *117*, 9906–9911.

Please cite this article in press as: Orr et al., Global Patterns and Drivers of Bee Distribution, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.053

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
DiscoverLife checklist	DiscoverLife, Polistes Foundation	https://www.discoverlife.org/mp/20q?guide=Apoidea_species&flags=HAS
iDigbio	Integrated Digitized Biocollections	http://www.idigbio.org 25 August 2018
SCAN	Symbiota Collections of Arthropods Network	https://www.scan-bugs.org/portal/index.php 28 August 2018
BISON	Biodiversity Information Serving our Nation	https://bison.usgs.gov 28 August 2018
GBIF	Global Biodiversity Information Facility	https://doi.org/10.15468/dl.dyyirp, https://doi.org/10.15468/dl.elno2c, https://doi.org/10.15468/dl.ig6jgr, https://doi.org/10.15468/dl.iqltv4, https://doi.org/10.15468/dl.tlttkn, https://doi.org/10.15468/dl.blpw69 (3 May 2018) https://doi.org/10.15468/dl.jzp4aa (16 August 2018).
ALA	Atlas of Living Australia	https://www.ala.org.au 25 August 2018
Table S1. Overlap and duplication between databases	Supplements	Table S1
Table S2. Synonym-matching file.	Supplements	Table S2
Table S3. Bee inventory data sources.	Supplements	Table S3
Data S1. Supplementary results tables.	Supplements	Data S1A-G
Data S2. Independent regression results.	Supplements	Data S2A-E
Data S3. Parameter estimates and descriptive statistics.	Supplements	Data S3A-C
Data S4. Sources for driver variables.	Supplements	Data S4
Software and Algorithms		
PAST	[40]	http://priede.bf.lu.lv/ftp/pub/TIS/datu_analiize/PAST/2.17c/download.html
Teraplot (3D graphs)	Kylebank Software Ltd.	https://www.teraplot.com
Maxent	American Museum of Natural History	https://biodiversityinformatics.amnh.org/open_source/maxent/
SAM: Spatial Analysis for Macroecology	[41]	http://www.ecoevol.ufg.br/sam/
GEODA	Lixun910	http://geodacenter.github.io/download.html
iNext	[42]	http://chao.stat.nthu.edu.tw/wordpress/software_download/
ArcGIS	ESRI	www.arcgis.com

RESOURCE AVAILABILITY

Lead Contact

Further information and requests on methods can be directed to Dr. Alice C. Hughes ach_conservation2@hotmail.com

Materials Availability

Physical materials were not used within this study

Data and Code Availability

Data used in the study is available through links provided on the Key Resources table. Maps are available as figures in supplements and main figures, and country level data is available as supplementary excel tables. Code is available through the iNext package.

Please cite this article in press as: Orr et al., Global Patterns and Drivers of Bee Distribution, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.053

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No physical experiments were made as part of the study

METHOD DETAILS

Data compilation and cleaning

Full methods are available in Methods S1. Checklist data[16] were compiled from various sources including taxonomic revisions, mostly cited in Michener,[32] and nearly all primary descriptions including for synonyms, although verifiable human observations and un-digitized museum records were also included (Methods S1A1). Checklists were compiled at present-day state- or country-level to avoid issues with historical political boundary changes (data are available online[16] in a matrix guide form at the country-level and species-by-species below the country-level). For areas with disparate data, neighboring units were merged (Russia, India, Indonesia, Philippines). In total, 168,618 unique species-area combinations were used. The public dataset included three major global data sources: Global Biodiversity Information Facility (GBIF-[43]), iDigBio (IDB-[44]), and Symbiota Collections of Arthropods Network (SCAN-[45]) and two regional sources: Atlas of Living Australia (ALA-[46]) and Biodiversity Information Serving Our Nation (BISON-[47]). These were selected as they are representative, well-known, and easy-to-access. Private sources were avoided as these largely serve developed countries and would only intensify biases, and analysis aims to utilize the public datasets most frequently used in large-scale analysis.

Data were initially uploaded into ArcMap 10.3, converted to point and exported for cleaning. Duplicates were removed and records listed only above species-level were deleted. Species were checked and corrected for spellings and synonyms using the checklist[16] and other sources (Table S2) to create a new synonym list for bees. Following this, species inappropriately placed in either the Eastern or Western Hemisphere (based on the checklist) were removed, though alien species were left as recorded. The number of species and samples were then calculated for each administrative unit based on the newly-cleaned databases, and the overlap (duplication) between databases calculated (Table S1). Administrative areas were at the highest resolution for which sufficient information existed, meaning state-level in many large countries, and country-level in small or little-known areas. Biogeographically complex areas (Philippines, Indonesia, etc.), and those with only regional checklists (Russia, India) were split into combined administrative areas given biogeography, sample size, and sample reliability to optimize quality and ecological relevance. Administrative areas are those at which checklists were compiled rather than statements of jurisdiction and do not represent political boundaries or ownership, most reflect state or country level boundaries. As a further quality metric, public data were compared to representative country checklists for 25 countries to check for further mismatches and provide an accuracy index. This check was only possible for better-known countries, so corrections were avoided to prevent bias.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mapping and data visualization

Collated records based on the cleaned public database and checklist produced global maps of richness of each administrative unit. The outputs of all these were compared via cartograms prior to further analysis, including exploring if the area of cartograms increased or decreased relative to actual size, as this denotes if a country has more or less species or samples than expected for its area. Cartograms develop a “global average” for the number of expected species per unit area, then deforms each unit (country/province) relative to if their value is greater or less than that global average.

Mapping and geospatial data analysis were conducted in ArcMap 10.3 with equal area projections. Basic analyses and data collation were conducted in Microsoft Excel, and regional metrics were calculated with ArcMap summary statistics. Cartograms were created using the cartogram toolbox for sampling and diversity based on the checklist and the databases.

Rarefaction analyses

Rarefaction analysis was used to estimate sampling completeness from databases. To understand this we assayed which countries could reach an asymptote on analysis based on a number of criteria to ensure sufficient coverage of that country, and prevent countries where a small area had been sampled intensively reaching an apparent asymptote purely due to poor or biased sampling effort. The iNEXT package[42] was used in R 3.4.4 to perform rarefaction analyses using the database at the country level (Methods S1B). To ensure representative sampling, countries were excluded from consideration when they failed the following *a priori* checks: sampling completeness > 0.80 (initial step, generated in iNext), a minimum size of 6,000km^2, a spatial coverage of 50% based on 25km2 grid cells, a minimum of four total grid cells with > 75% terrestrial cover (i.e., 75% of land-surface covered), a minimum of 250 total records, and plots clearly asymptoting. These criteria were developed to eliminate areas with low sampling coverage, or small islands as such datasets could falsely asymptote at low levels due to intense sampling of a small area. Asymptotes were assessed by eye by three individuals (ACH, DC, MCO) independently and scored, if only two individuals decided that asymptotes had been reached this was discussed before deriving the final list. The axes used for each graph to examine if an asymptote was reached was automatically scaled for each country to facilitate describing trends in regions with very different totals and to further standardize asymptote decisions.

Drivers analyses and mapping richness

In addition to highlighting patterns of bias, we used the checklist to determine the drivers of global bee richness, and patterns at a higher resolutions in addition to conducting regional analyses. Thirty-one ecophysiologically-relevant variables were initially selected as factors which could play a significant role in determining species distributions including geographic and climatic factors (Data S4). Explanations behind variable choice, and generation of datalayers is available in Methods S1A4.

We calculated the mean and standard deviation for each of 31 variables (to give 62 variables - Data S4) within each administrative area using the zonal statistics tool in Arcmap, providing the mean and standard deviation of each parameter within each area. Independent and group stepwise regression between the 62 biologically-relevant variables and richness based on the checklist returned a final set of 18 most-important parameters based on AIC and R (Methods S1C).

The software programs PAST[40] and SAM[41] were used to run standard linear regressions, and this was compared to an Ordinary least-squares regression run in ArcGIS and these were then used to generate a model of how these variables influence bee species richness globally and regionally, by running this analysis both for the world as a whole, and for each of the three regions independently (New world, Eurasia-Africa, Australia). Global-level centroid averages should not show autocorrelation (verified using Morans I) the models above could be used for assessing variable importance, but could have issues with predicting richness. Thus, an additional model was run in Geoda where model output is provided and collinearity could be maintained below acceptable levels of 30 (Data S2,S5, Methods S1A5) (Anselin 2006).

To examine the diversity patterns of bees in relation to global climatic zones (akin to Köppen-geiger zones, but based on more ecophysiologally relevant variables for this study), all 31 variables were divided into three categories: energy (or directly available resources), precipitation and temperature. Within each of these an sPCA was run to explore environmental variation, and for each the first layer (equivalent to the 1st axis) was kept. sPCAs provide a spatial approach to collapsing environmental variation down to the minimum number of layers (represented as axes). Isocluster analysis was then used to identify climate zones for each axis by using the sPCA to cluster areas with similar environmental conditions into a single zone. The ''Energy'' isocluster analysis reconstructed Mediterranean, xeric, and sometimes temperate regions (Methods S1D).

The conditions on the three layers were then extracted for all species locality data from the public database to provide a measure of each axis for each record, which was then averaged for each species using the summary statistics tool. The recorded richness on the checklist was compared to the average sPCA value for each administrative unit. Then both database individual species records and the checklist richness compared to each sPCA axis were plotted in turn on 3D graphs to explore how species distributions and richness varied relative to the conditions present.

MaxEnt was then used to predict and map global bee richness based on the first layer of each of the three sPCA categories (minimizing correlation and redundancy between variables), combined with richness counts (rather than species) of areas (10km grids) with a minimum of 50 samples, and split into divisions of ten (i.e., 1-10 species, 11-20 etc). This was because sample-size may be too low for smaller areas, a 10km resolution for data aggregation represented a balance between sufficient data and environmental heterogeneity (which would increase with area size), in addition differences in sampling intensity could not be accounted for, so this sort of approach is likely to be more representative. Models for each diversity level were run with five iterations and an average taken, then reclassified using the ten percentile training presence threshold as a baseline for unsuitable. Above this baseline value, the probability of occurrence was split into 10 divisions to match the original values as similar to probability of occurrence, higher value areas likely supported more species. Model outputs were then mosaiced to give the maximum number of species the area was suitable for based on all models together. Initially the world was split into regions and a model run for each region (Table S3), but the outputs failed to capture richness in some species-rich, under-sampled regions (Figure S5), so models were rerun based on analysis from the Americas and reprojected to the rest of the world. Models were then compared to the checklist patterns and verified by experts to assess how well they matched known patterns of bee richness (listed in supplement). Richness models have become a popular way to assess relationship between richness and environmental parameters and to look at richness and turnover even over poorly sampled areas, rather than stacking individual species models.[48,49]

Regional diversity patterns

Higher resolution data in the New World enabled more sophisticated analysis for the region based on the databases. We created minimum convex polygons (MCPs) using the minimum bounding geometry tool based on database point records for each species to create a convex hull around all localities each species was recorded at. These were then trimmed by a polygon of the landmass of the New world. Line density tool was used at a 10km resolution to map the co-occurrence of species range boundaries generated using the MCPs, once the minimum convex polygons had been converted to polylines.

Richness was generated from MCPs by converting to rasters and giving them a value of one, then using the mosaic tool to sum values. A spatial Principal Components Analysis (sPCA) was used on these species occurrence rasters (once reclassed to show unsuitable regions as zero, and removing ranges of species limited to few, tightly clustered sites (i.e., all locations within a total area of under five km) where more sampling had occurred and may otherwise inflate perceived richness and endemism at well-sampled sites) to explore compositional changes, as evidenced by the co-occurrence of different species based on the suitable and unsuitable rasters. The sPCA shows how the composition of species present (based on the MCPs) varies over space, thus areas with more different colors show a more dissimilar community makeup.

Turnover was calculated and visualized by calculating the latitude and longitude of each species point records to the nearest integer value then used the summary statistics tool to calculate minimum and maximum range limits for each species, and calculating how many species showed these limits at each latitude, then showing this graphically to assay if many species showed similar geographic limits to their ranges, and complemented by Figure 5C which shows that major turnover occurs along the coast and at continental coastlines, in addition to at the tropical-subtropical intersection (Figure S3A).

The outputs of these steps provide an idea of the community makeup and turnover across the Americas on a continuous basis. As sampling was uneven, these steps are only suggestive of patterns, but still provide insights into patterns of richness and turnover across the region.

Current Biology, Volume *31*

Supplemental Information

Global Patterns and Drivers of Bee Distribution

Michael C. Orr, Alice C. Hughes, Douglas Chesters, John Pickering, Chao-Dong Zhu, and John S. Ascher

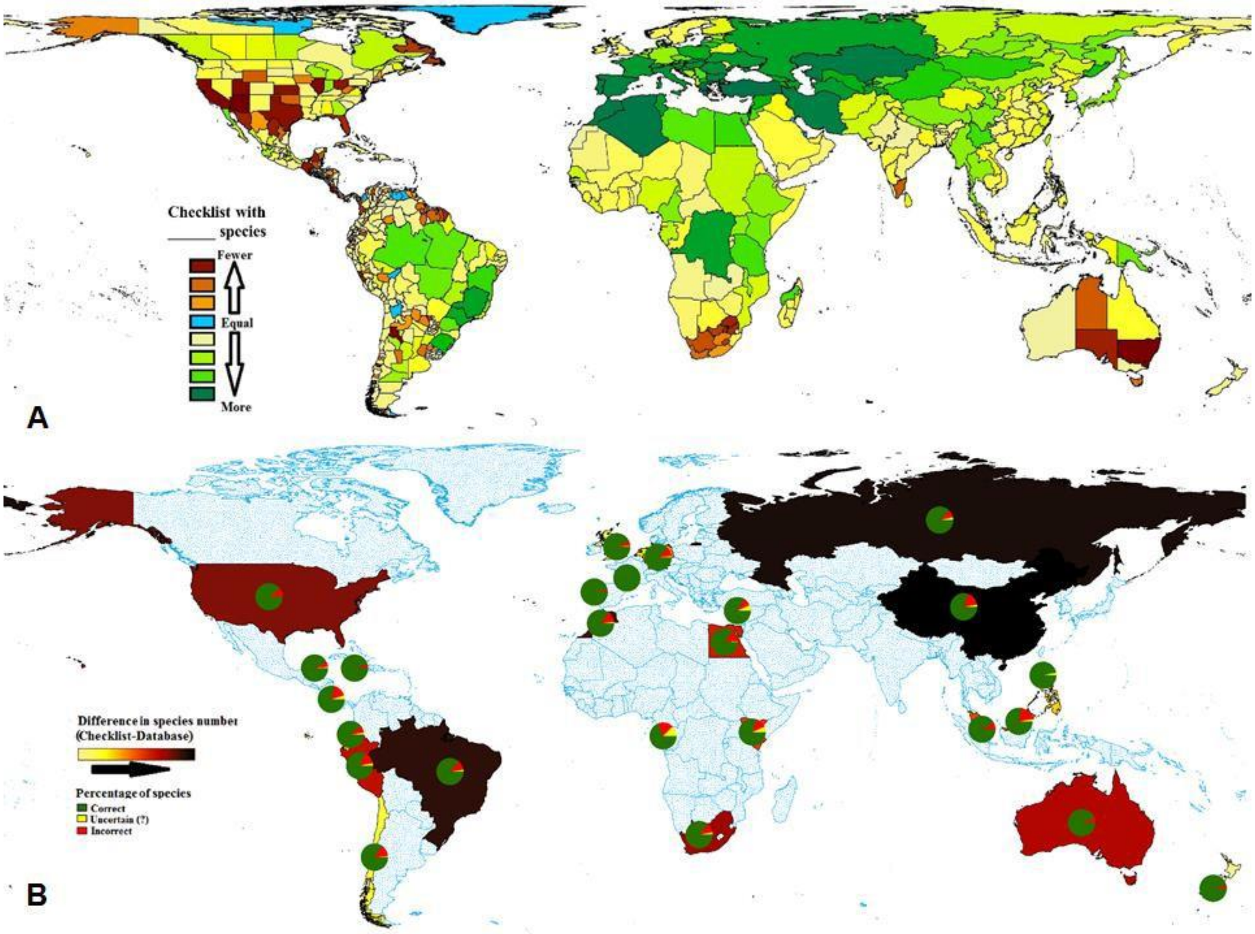

Figure S1. Additional error checks. Related to Figure 3. A. Differences in richness between checklist and public databases. Units with more species in the checklist than the public database are given in darker green, those in blue have the same richness, and darker brown denotes less species in the checklist than in the public database. Note that this is prior to country-level error checking; all countries have fewer species recorded following that step (B). B. Country-level species checks. 25 countries were checked and are color-coded from light to dark based on lower to greater deficits in number of species reported by the public database (after these checks were made) vs the checklist, with darker countries showing higher under-representation in the public databases. Unchecked countries are stippled out. Pie-charts indicate the proportion of correct (green), uncertain (yellow), and definitively incorrect (red) species listed by the public database, based on

individual crosschecking using the checklist (correct) and possible species not yet included in the checklist (uncertain). All countries contain fewer species than the checklist following these checks, with a minimum of 17 less in the public database and a maximum of 1137. Additional rarefaction checks are available in Figure S2.

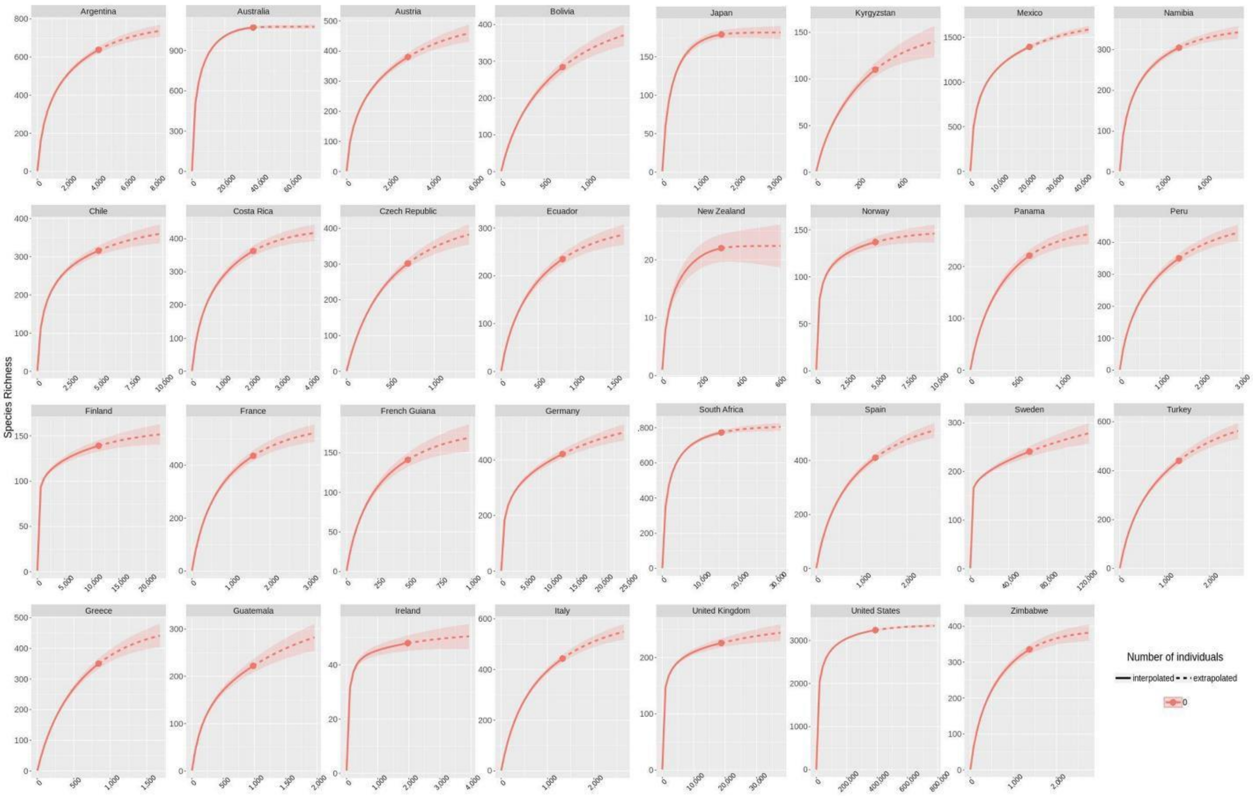

Figure S2. Rarefaction curves for the countries passing the criteria on sampling. Related to Figure 3. 31 countries in total. The following countries were judged to asymptote: Australia, Finland, and Ireland, Japan, New Zealand, Norway, South Africa, United Kingdom, and the United States. See also Figure 3 for spatial coverage.

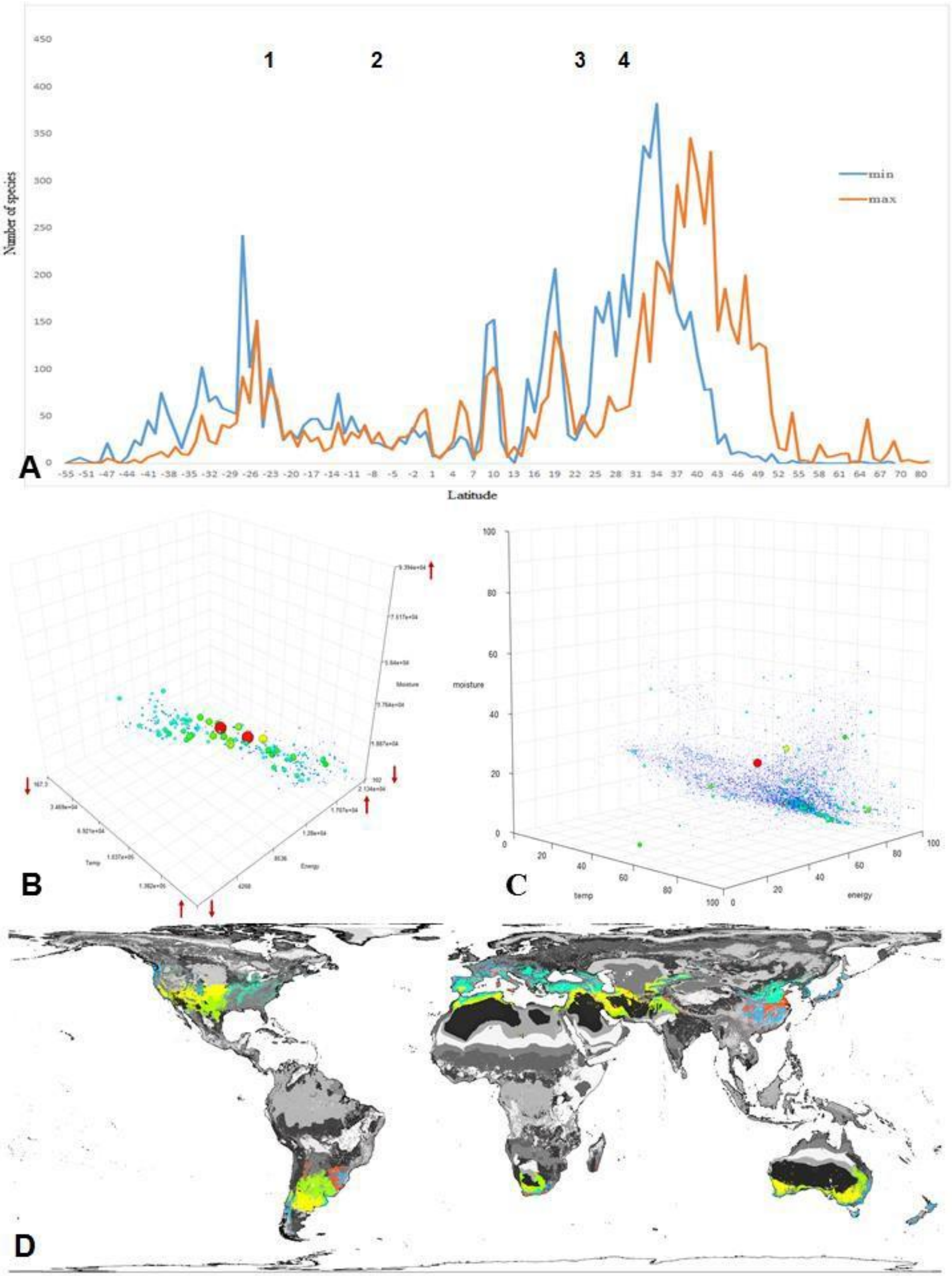
A
1
2
3
4
Number of species
450
400
350
300
250
200
150
100
50
0
min
max
-55 -51 -47 -44 -41 -38 -35 -32 -29 -26 -23 -20 -17 -14 -11 -8 -5 -2 1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 80
Latitude
B
9.394e+04
7.517e+04
5.64e+04
Moisture
3.764e+04
1.887e+04
102
2.134e+04
1.707e+04
1.28e+04
8536
Energy
4268
167.3
3.469e+04
6.921e+04
1.037e+05
Temp
1.382e+05
C
100
80
60
moisture
40
20
0
0
20
40
60
80
100
temp
0
20
40
60
80
100
energy
D

Figure S3. Environmental parameterization and limits. See also Figures 4 and 6. A. Maximum and minimum latitudes of occurrence for species in the Americas. The number of species sharing a maximum/minimum latitude of occurrence is denoted on the Y axis. Significant features are indicated by 1-4. 1: sub- tropical-tropical division in Latin America, 2- Northern edge of South America, 3 Tropics- subtropics in North America 4. Gulf of Mexico. High peaks are clear on the tropical-subtropical interface. B. This graph shows the position of each region in environmental space. Richness goes from dark blue (low richness)-green-yellow-red with increasing richness, and from small to large. Species-rich areas can be seen to occupy a subset of conditions and cluster in environmental space: low moisture, high energy, and variable but seemingly intermediate temperature. Red arrows indicate axis value directionality. C. This graphic shows the mean conditions based on each of the three axes for each species. Condition axes were rescaled to a percentage of the maximum to enable clearer interpretation (0-100%). Species sharing identical conditions were calculated and are represented by larger and more brightly-colored icons (low to high, blue-green-yellow-red). Though species occupy a wide range of conditions, most species only occupy a subset of the total climate space available. The characteristics described here match those of arid and xeric regions, with a steady cline towards temperate environments as highlighted by the regions highlighted in Figure S3d, methods are provided in Methods S1A6.

Red arrows indicate axis value directionality. D. Isocluster of climate zones for “energy” sPCA. Colored areas show different types of Mediterranean and xeric climate. Different colors denote slight changes in conditions within such climates, and show that bee richness hotspots often have similar conditions based on energy related factors.

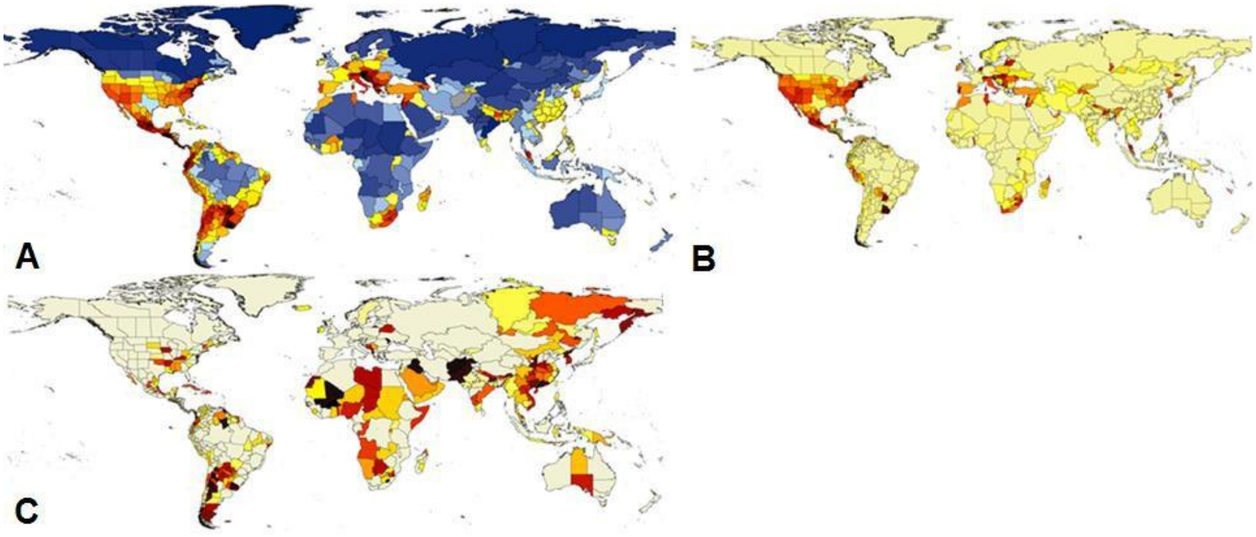

Figure S4. Reprojected species richness by administrative areas, various representations. Related to Figure 6. A. Actual area compared to estimated area based on richness from the cartogram. Dark-pale blue indicates the area has fewer than expected species for its area, with the darkest blue areas having particularly low richness. Yellow-red-black increasingly indicates more species than expected for its area, with the darker countries having the highest number of species per-unit-area. Grey countries contain the "average" number of species per-unit-area. B. Number of species per unit area based on simple relationship between number of species and area (not using projected increase, darker hues indicate higher richness per unit area). C. Difference between projected and checklist richness based on driver analysis with more intense hues (Yellow-Orange-Red-Black) indicating higher projected richness relative to known richness.

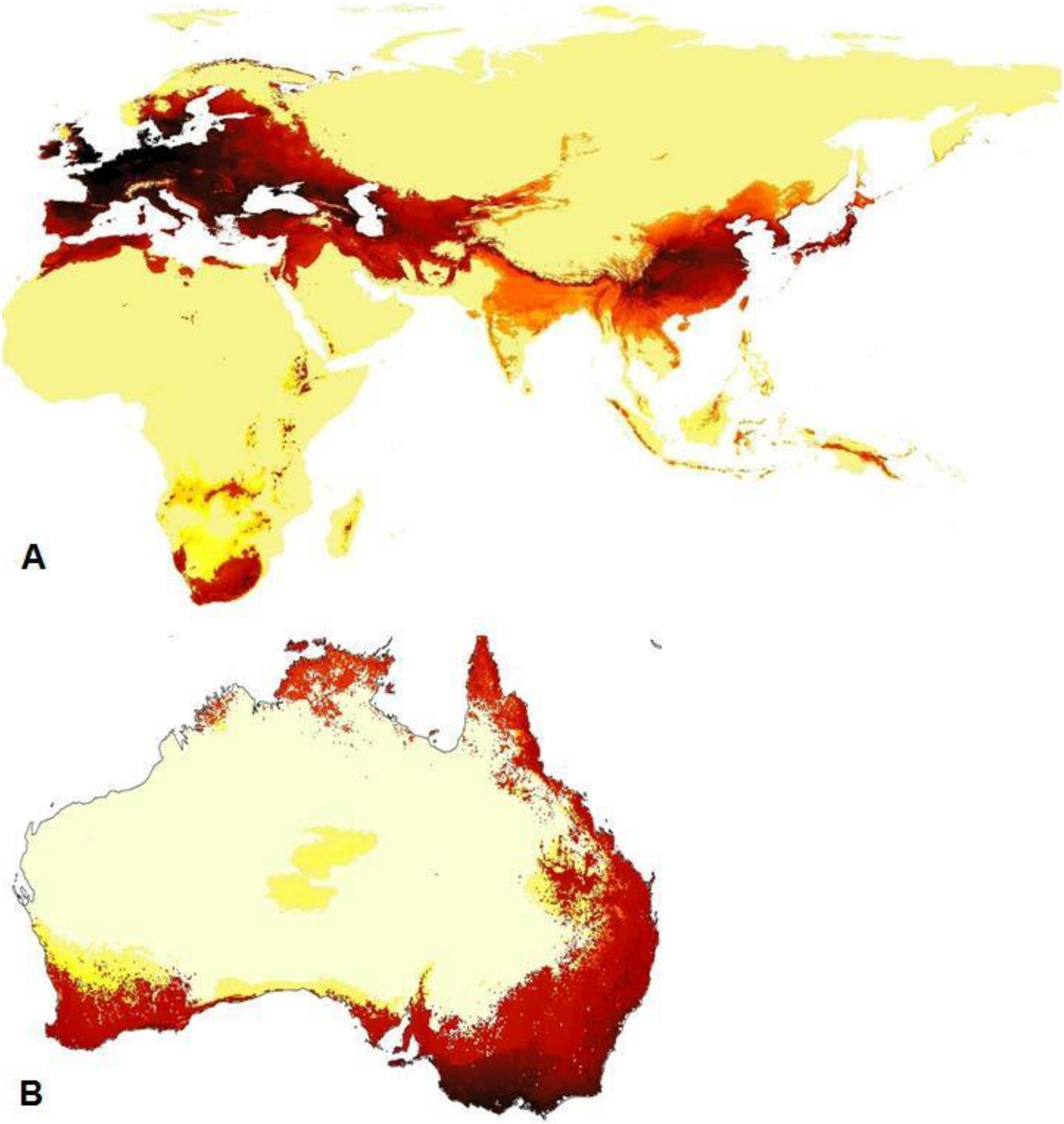

Figure S5. Predicted richness for the Old World using Old World point data. Related to Figure 6. A. Predicted richness of Eurasia and Africa using Old World point data. The disproportionately high sampling in northern Europe heavily inflated richness there while a lack of sampling throughout most areas made realistic modeling impossible with these data. B. Predicted richness of Australia

using Australian point data. This model for Australia clearly underperformed due to intense sampling biases..

			Duplicates		
DB1	DB2	Combined	removed	%unique	%loss
Bison	ALA	237569	233865	98	2
GBIF	ALA	411298	344511	84	16
IDB	ALA	243029	233291	96	4
SCAN	ALA	144975	121866	84	16
GBIF	BISON	595409	538833	90	10
IDB	BISON	427139	426293	100	0
SCAN	BISON	329085	309240	94	6
IDB	GBIF	600869	548727	91	9
SCAN	GBIF	502813	396822	79	21
SCAN	IDB	334546	237419	71	29

Table S1. Duplicate records between each of the public databases. Related to Figure 3, Table 1.

Supplemental References

Note that inventory references used in the initial richness models (to improve Old World performance) are given in Table S5.

S1. Michener, C.D. (2007). The Bees of the World (Johns Hopkins University Press, Baltimore, ed. 2).

S2. Nieto, A. et al. (2014). European red list of bees (International Union for Conservation of Nature, Luxembourg, Publication Office of the European Union).

S3. Leijs, R., Dorey, J., & Hogendoorn, K. (2018). Twenty six new species of *Leioproctus* (*Colletellus*): Australian Neopasiphaeinae, all but one with two submarginal cells (Hymenoptera, Colletidae, *Leioproctus*). ZooKeys *811*, 109–168 (2018).

S4. Michener, C.D. (1979). Biogeography of the bees. Ann. MO Bot. Gard. *66*, 277–347.

S5. Ascher, J.S. & Pickering, J. (2018). Discover Life bee species guide and world checklist, (Hymenoptera: Apoidea: Anthophila), Draft 51, https://www.discoverlife.org/mp/20q?guide=Apoidea_species
Date: 8 November 2018

S6. Danforth, B.N., Minckley, R.L., Neff, J.L., & Fawcett, F. (2019). The solitary bees: biology, evolution, conservation (Princeton University Press, Princeton).

S7. Ascher, J.S. (2016). Collaborative databasing of North American bee collections within a global informatics network project. iDigBio Darwin Core Archive Recordset. https://www.idigbio.org/portal/recordsets/8919571f-205a-4aed-b9f2-96ccd0108e4c [369, 654 bee specimen records derived from the AMNH-owned Arthropod Easy Capture database. https://sourceforge.net/p/arthropodeasy/wiki/Home/]

S8. Hughes, A.C., Satasook, C., Bates, P.J., Bumrungsri, S., & Jones, G. (2011). Explaining the causes of the zoogeographic transition around the Isthmus of Kra: using bats as a case study. J. Biogeo. *38*(12), 2362–2372.

S9. Clarke, A., & Gaston, K.J. (2006). Climate, energy and diversity. Proc. Biol. Sci. *273*(1599), 2257–2266.

S10. Vieira, T.B., et al. (2018). A multiple hypothesis approach to explain species richness patterns in neotropical stream-dweller fish communities. PLoS ONE *13*(9), e0204114.

S11. McCabe, L.M., Colella, E., Chesshire, P., Smith, D., & Cobb, N.S. (2019). The transition from bee-to-fly dominated communities with increasing elevation and greater forest canopy cover. PLoS ONE *14*(6), e0217198.

S12. Ellis, E.C., Antill, E.C., & Kreft, H. (2012). All is not loss: plant biodiversity in the Anthropocene. PLoS ONE *7*, e30535.

S13. Burnham, K.P. & Anderson, D.R. (2001). A practical information–theoretic approach. Model selection and multi-model inference, New York, NY: Springer 2nd ed.

S14. Anselin, L. (2006). Exploring spatial data with Geoda: A workbook, spatial analysis laboratory department of geography (University of Illinois, center for spatially Integrated social science).

S15. Variables for drivers analysis (see Data S4): ANNPET, ARIDITYT, CMI, CONTINEN, EMBERRQ, GDD0, GDD5, PETSEA, thermInd.

http://envirem.github.io/

Date: March 2019

S16. Variables for drivers analysis (see Data S4): PET_HE, AI.

https://cgiarcsi.community/data/global-aridity-and-pet-database/

Date: March 2019

S17. Variables for drivers analysis (see Data S4): ME_ND_NF, SD_NV_NF, ME_NDVI, STD_NDVI.

https://land.copernicus.vgt.vito.be/PDF/datapool/Vegetation/Indicators/NDVI_1km_V2/

Date: March 2019

S18. Variables for drivers analysis (see Data S4): TEMP2M, DNI.

https://globalsolaratlas.info/downloads/world

Date: March 2019

S19. Variables for drivers analysis (see Data S4): PRICH_NF, st_prich_nf.

http://ecotope.org/files/used_planet/ellis_etal_2013_dataset.zip

Date: March 2019

S20. Variables for drivers analysis (see Data S4): BIO1-6, BIO12-15, SRAD, WIND, VAPR.

http://worldclim.org/version2

Date: March 2019.

S21. Variable for drivers analysis (see Data S4): AET.

https://figshare.com/articles/Global_High-Resolution_Soil-Water_Balance/7707605/3

Date: March 2019.

S22. Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. Journal of Geophysical Research: Biogeosciences *116*(G4),

S23. Layer used for variable construction for drivers analysis (see Data S4): Tree density.

https://elischolar.library.yale.edu/yale_fes_data/1/

Date: March 2019